Physiological time-series analysis: what does regularity quantify?

STEVEN M. PINCUS AND ARY L. GOLDBERGER
Cardiovascular Division, Department of Medicine, Beth Israel Hospital and Harvard Medical School, Boston, Massachusetts 02215

Pincus, Steven M., and Ary L. Goldberger. Physiological time-series analysis: what does regularity quantify? *Am. J. Physiol.* 266 (*Heart Circ. Physiol.* 35): H1643–H1656, 1994.—Approximate entropy (ApEn) is a recently developed statistic quantifying regularity and complexity that appears to have potential application to a wide variety of physiological and clinical time-series data. The focus here is to provide a better understanding of ApEn to facilitate its proper utilization, application, and interpretation. After giving the formal mathematical description of ApEn, we provide a multistep description of the algorithm as applied to two contrasting clinical heart rate data sets. We discuss algorithm implementation and interpretation and introduce a general mathematical hypothesis of the dynamics of a wide class of diseases, indicating the utility of ApEn to test this hypothesis. We indicate the relationship of ApEn to variability measures, the Fourier spectrum, and algorithms motivated by study of chaotic dynamics. We discuss further mathematical properties of ApEn, including the choice of input parameters, statistical issues, and modeling considerations, and we conclude with a section on caveats to ensure correct ApEn utilization.

approximate entropy; complexity; chaos; stochastic processes; nonlinear dynamics; heart rate variability

PHYSIOLOGISTS AND CLINICIANS are confronted frequently with the problem of comparing time-series data such as the two heart rate (HR) tracings seen in Fig. 1, both obtained on 4-mo-old infants during quiet sleep (36). Figure 1*A* is from an infant who had an aborted sudden infant death syndrome (SIDS) episode[1] ~1 wk before the recording, and Fig. 1*B* is from a healthy infant. The overall variabilities [standard deviations (SDs)] of these two tracings are approximately equal, and whereas the aborted-SIDS infant has a somewhat higher mean HR, both are well within the normal range. Yet the tracing in Fig. 1*A* appears to be more regular (less complex) than the tracing in Fig. 1*B*. Therefore, we ask, *1*) How do we quantify this apparent difference in regularity? *2*) Does this measure give a significantly lower value for the aborted-SIDS infant compared with a range of normal infant values? *3*) How do the inherent limitations posed by a single epoch of quiet sleep (~1,000 points of HR data, with some system noise present) affect statistical analysis? and *4*) What is a possible physiological basis for greater regularity (decreased complexity) under pathological conditions?

Over the past three years, a new mathematical approach and formula termed approximate entropy (ApEn) has been introduced as a quantification of regularity in data, motivated by the four questions above. Mathematically, ApEn is part of a general theoretical development, as the natural information theoretical parameter for an approximating Markov Chain to a process (35). In applications to a range of medical settings, findings (20, 21, 34, 36, 37, 40) have associated sickness and aging with significantly decreased ApEn values, consistent with our general hypothesis associating compromised physiology in many systems with more regular, patterned sinus rhythm HR tracings, and normative physiology with greater irregularity (randomness, complexity). Findings implicitly associating greater regularity with compromised physiological status have been found elsewhere, including in spectral analysis of HR in preterm babies (1) and eventual SIDS infants (23) and in adult sudden death syndrome victims (14) as well as in analysis of electrocardiographic waveforms during fatal ventricular tachyarrhythmias (15).[2] These findings have produced qualitatively interesting results but lack a clear-cut statistic that summarizes the frequency spectra or underlying system structure.

We thus see that a suitable quantification of ensemble regularity vs. randomness in (HR and other) data has the potential to provide insights in a wide range of clinical settings, both diagnostic and predictive. However, there are always technical dangers in applying new mathematical developments: the application may be out of context (with no statistical validity); researchers may apply a newly developed tool to establish a distinction that could have been made with previously well-established techniques; or when a new formula or algorithm is employed as a "black box," without a

[1] An aborted-SIDS infant (also described as an infant experiencing an apparent life-threatening event) is defined as an apparently healthy infant who had experienced an episode of unexplained apnea with cyanosis or pallor requiring mouth-to-mouth resuscitation or vigorous physical stimulation for revival.

[2] We note, of course, that not all disease processes are associated with greater regularity. The HR time series for the ventricular response to atrial fibrillation in humans, in the absence of atrioventricular node disease or drug effects, apparently resembles white noise (13), both in the time and frequency domain, and thus will exhibit larger ApEn values than seen in normal sinus rhythm.

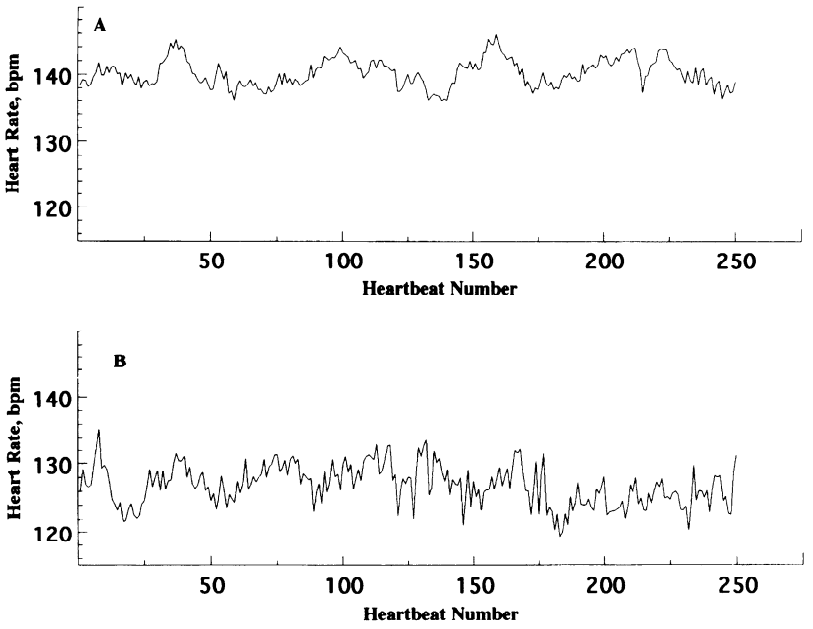

Fig. 1. Comparison of infant quiet sleep heart rate (HR) tracings with similar overall variability (Var; SD). *A*: aborted-sudden infant death syndrome (-SIDS) infant; Var = 2.49 beats/min (bpm), approximate entropy (ApEn) = 0.742. *B*: normal infant; Var = 2.61 bpm, ApEn = 1.457.

suitably developed mathematical intuition, counterintuitive findings may result, and the user would be unsure as to the "cause" of the confounding answers. The purpose of this tutorial is to provide a detailed understanding of ApEn and the notion of regularity so analysts of physiological time-series data such as HR can avoid these pitfalls.

QUANTIFICATION OF REGULARITY

Definition of ApEn

Two input parameters, m and r, must be fixed to compute ApEn: m is the "length" of compared runs, and r is effectively a filter. For fixed m and r, we define both the parameter (the true number, defined as a limit) ApEn(m,r), and the statistical estimate ApEn(m,r,N) given N data points $u(1)$, $u(2)$, ..., $u(N)$. To develop an impression of what ApEn quantifies, the reader need not initially be overly concerned with the distinction between parameter and estimate; in CHOICE OF INPUT PARAMETERS AND STATISTICAL AND MODEL-RELATED ISSUES we concern ourselves with these distinctions directly. Given N data points $\{u(i)\}$, form vector sequences $\mathbf{x}(1)$ through $\mathbf{x}(N - m + 1)$, defined by $\mathbf{x}(i) = [u(i), ..., u(i + m - 1)]$. These vectors represent m consecutive u values, commencing with the ith point. Define the distance $d[\mathbf{x}(i),\mathbf{x}(j)]$ between vectors $\mathbf{x}(i)$ and $\mathbf{x}(j)$ as the maximum difference in their respective scalar components. Use the sequence $\mathbf{x}(1)$, $\mathbf{x}(2)$, ..., $\mathbf{x}(N - m + 1)$ to construct, for each $i \leq N - m + 1$, $C_i^m(r)$ = (no. of $j \leq N - m + 1$ such that $d[\mathbf{x}(i),\mathbf{x}(j)] \leq r$) / $(N - m + 1)$. The $C_i^m(r)$ values measure within a tolerance r the regularity, or frequency, of patterns similar to a given pattern of window length m. Define $\Phi^m(r) = (N - m + 1)^{-1} \sum_{i=1}^{N-m+1} \ln C_i^m(r)$, where ln is the natural logarithm, and then define the parameter ApEn(m,r) = $\lim_{N\to\infty}[\Phi^m(r) - \Phi^{m+1}(r)]$.

For virtually all reasonable processes, this definition is well defined (a unique limit exists with probability one). Given N data points, we estimate this parameter by defining the statistic ApEn(m,r,N) = $\Phi^m(r) - \Phi^{m+1}(r)$. ApEn measures the (logarithmic) likelihood that runs of patterns that are close for m observations remain close on next incremental comparisons. Greater likelihood of remaining close, regularity, produces smaller ApEn values, and conversely.

On unravelling definitions we deduce[3] the essential observation that

$$\begin{aligned} -\text{ApEn} = \Phi^{m+1}(r) - \Phi^m(r) = \text{average over} \\ i \text{ of ln [conditional probability that} \\ |u(j + m) - u(i + m)| \leq r, \text{ given that} \\ |u(j + k) - u(i + k)| \leq r \text{ for } k = 0, 1, \ldots, m - 1] \end{aligned} \quad (1)$$

What Does Entropy Quantify?

To provide a more intuitive, physiological understanding of this definition, we develop a multistep description of the algorithm with accompanying figures. We illustrate this procedure using the time series shown in Fig. 1*A*, the HR tracing for an aborted-SIDS infant. ApEn is calculated with $m = 2$ and $r = 0.6$ beats/min (36), consistent with guidelines indicated in IMPLEMENTATION AND INTERPRETATION.

The flow here is that given a time series of N (typically evenly sampled) data points, we form an associated time

[3] $-\text{ApEn} = \Phi^{m+1}(r) - \Phi^m(r) = [(N - m)^{-1} \sum_{i=1}^{N-m} \ln C_i^{m+1}(r)] - [(N - m + 1)^{-1} \sum_{i=1}^{N-m+1} \ln C_i^m(r)]$, which approximately equals $(N - m)^{-1} [\sum_{i=1}^{N-m} [\ln C_i^{m+1}(r) - \ln C_i^m(r)]]$ with equality in the limit as $N \to \infty$. Because logarithms satisfy a basic identity $\ln a - \ln b = \ln (a/b)$ for all positive a and b, this latter expression equals $(N - m)^{-1} \{\sum_{i=1}^{N-m} \ln[C_i^{m+1}(r)/C_i^m(r)]\}$, which equals the average over i of $\ln [C_i^{m+1}(r)/C_i^m(r)]$. This last parenthetical expression is readily seen to be the conditional probability indicated in *Eq. 1*. Thus ApEn can be calculated either as a difference (between Φ^m and Φ^{m+1}) or as an average of quotients or ratios (given by *Eq. 1*).

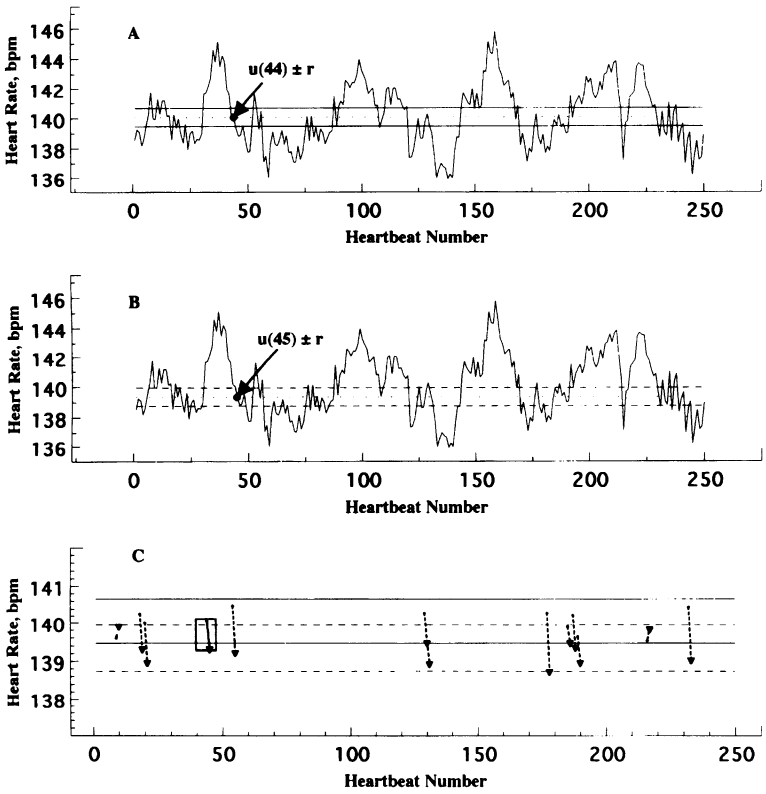

Fig. 2. Identification of all length 2 vectors $\mathbf{x}(j) = [u(j), u(j + 1)]$ that are component-wise close to $\mathbf{x}(44) = [u(44), u(45)]$, for HR tracing shown in Fig. 1*A*. These form a base set of comparison vectors from which conditional probabilities are calculated. *A*: solid lines, 0.6 above and 0.6 below dotted line, which passes through $u(44)$. Points $u(j)$ that are close to (within a band of width $r = 0.6$) the value of $u(44)$ are those points between the solid lines. *B*: dashed lines are again 0.6 above and below the dotted line, which now passes through $u(45)$. Points $u(j + 1)$ between these dashed lines are those that are close to $u(45)$. C: all length 2 vectors $\mathbf{x}(j) = [u(j), u(j + 1)]$ that are component-wise close to $\mathbf{x}(44) = [u(44), u(45)]$. Tail of each vector $[u(j)$ point] is required to be between solid lines (given in *A*), and tip of each vector $[u(j + 1)$ point] is required to be between dashed lines (given in *B*).

series of $N - m + 1$ vectors, each vector consisting of m consecutive points. For example, with $m = 2$ here, the first three vectors are $[u(1), u(2)]$, $[u(2), u(3)]$, and $[u(3), u(4)]$. Each vector serves, in turn, as a template vector for comparison with all other vectors in the time series, toward the determination of a conditional probability associated with this vector. The conditional probability (estimate) calculation consists of first obtaining a set of conditioning vectors close to the template vector and then determining the fraction of instances in which the next point after the conditioning vector is close to the value of the point after the template vector. Finally, ApEn aggregates these conditional probabilities into an ensemble measure of regularity.

Step I. We begin by focusing on the length 2 ($m = 2$) vector $[u(44), u(45)]$, denoted $\mathbf{x}(44)$. This vector represents two contiguous observations that graphically corresponds to part of a downslope of the waveform formed by connecting consecutive points in the time series. We begin with this vector, rather than $[u(1), u(2)]$ for purely representational reasons; in *step IV*, we see that conditional probabilities are calculated for all (length 2) vectors, then log averaged.

Step II. Identify all length 2 vectors $\mathbf{x}(j) = [u(j), u(j + 1)]$ that are component-wise close to $\mathbf{x}(44) = [u(44), u(45)]$, as illustrated in Fig. 2. As defined in *Definition of ApEn*, $\mathbf{x}(j)$ is close to $\mathbf{x}(44)$ if both $|u(j) - u(44)| \leq 0.6$ and $|u(j + 1) - u(45)| \leq 0.6$. There are 12 such vectors [excluding $\mathbf{x}(44)$ itself].

Step III (*central calculation*). Compute the conditional probability that $u(j + 2)$ is close to $u(46)$ ($< r = 0.6$ apart), given that the conditioning vector $\mathbf{x}(j) = [u(j), u(j + 1)]$ is component-wise close to $\mathbf{x}(44) = [u(44), u(45)]$, illustrated in Fig. 3. The conditional probability here is a ratio A/B, where A is the number of instances that $u(j + 2)$ is close to $u(46)$ and $\mathbf{x}(j)$ is close to $\mathbf{x}(44)$ and B is the number of instances when $\mathbf{x}(j)$ is close to $\mathbf{x}(44)$.

Visually, we evaluate candidate $u(j + 2)$ values only among those js (vectors) indicated in Fig. 2C. We see that 5 of 12 candidate values of $u(j + 2)$ fall in the indicated band, thus the conditional probability that runs of patterns that are close to the two observations commencing at $j = 44$ remain close on next incremental comparison = 5/12.

Step IV. Repeat *steps I–III* for each length 2 vector $\mathbf{x}(i)$, calculating a conditional probability. Calculate the average of the logarithm of these conditional probabilities; ApEn is defined as the negative of this value (to ensure a positive number).

The conditional probabilities calculated in *step III* are by definition between 0 and 1. The reader should observe that pronounced regularity, with recognizable patterns that repeat, will produce conditional probabilities closer to 1, and thus their logarithms (negative numbers) will be closer to 0. This will result in smaller ApEn values. Conversely, apparently random behavior will produce conditional probabilities closer to 0, since the closeness of the conditioning vectors will have little bearing on the closeness of subsequent comparisons. Taking logarithms here will thus produce relatively large negative numbers and, hence, larger ApEn values. The opposing extremes are perfectly regular sequences, (e.g., sinusoidal behavior, very low ApEn) and independent sequential processes (e.g., random walk, very large ApEn).

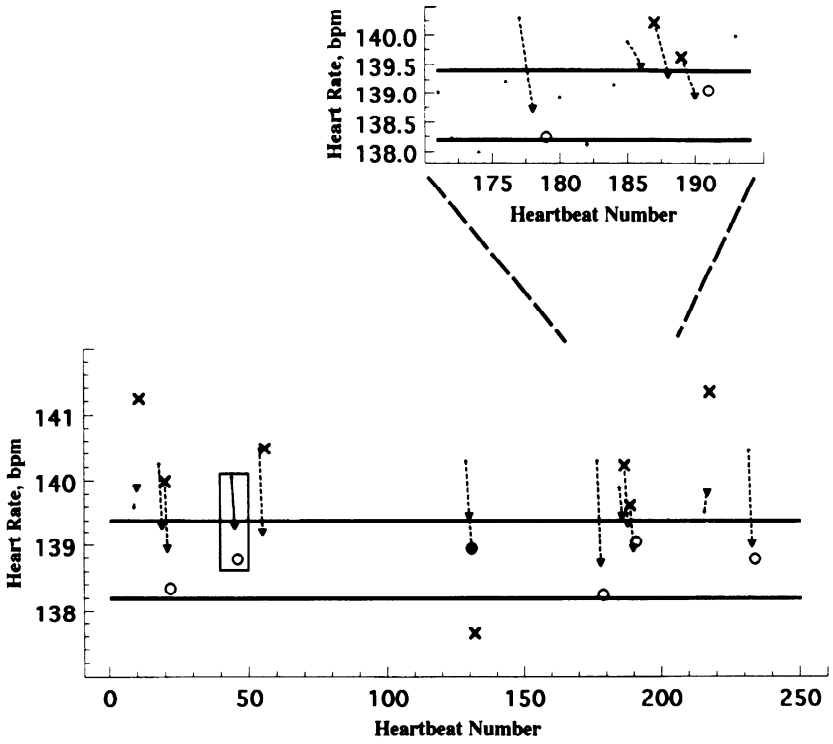

Fig. 3. Determination of those points $u(j + 2)$ that are close to $u(46)$, given that the conditioning vector $\mathbf{x}(j) = [u(j), u(j + 1)]$ is close to $\mathbf{x}(44) = [u(44), u(45)]$. ○, Close points, falling between solid lines width 0.6 above and below the value of $u(46)$; ×, not close points. Rectangle encloses template vector $\mathbf{x}(44)$ and "next point" $u(46)$. Conditioning vectors are those shown in Fig. 2*C*. Excluding reference point $u(46)$, 5 candidate values of $u(j + 2)$ are close to $u(46)$, whereas 7 candidate values of $u(j + 2)$ are not close to $u(46)$. We thus form the conditional probability of the likelihood that runs of patterns close to the 2 observations commencing at $j = 44$ ($5 + 7 = 12$ conditioning vectors) remain close on next incremental comparison (5 successful points), as $5/12$. *Top*: close-up view near $u(180)$, including entire time series.

Clinically, relatively low ApEn values (e.g., for HR) appear to correlate with pathology. For example, in Fig. 1 the ApEn value of the aborted-SIDS infant (0.742) was significantly smaller than any of 45 normal infant ApEn values (36). To illustrate this point visually, we consider Fig. 4, generated in the same manner as Fig. 3 with the difference that the baseline time series is that of Fig. 1*B*, from the normal infant. We again calculate the conditional probability that $u(j + 2)$ is close to (within 0.6 of) $u(46)$, given that $\mathbf{x}(j)$ is component-wise close to $\mathbf{x}(44)$. Here, only 1 of 11 candidate values of $u(j + 2)$ is close to $u(46)$, yielding a conditional probability of $1/11$. The negative logarithm of $1/11$ (2.398) is relatively large and larger than $-\log 5/12$ (0.875), as computed for the aborted-SIDS infant, indicating greater apparent randomness from this template vector and contributing to an overall larger normal infant ApEn value.

Figure 4, *top*, provides insight to the small number of candidate u values that land within the band of width 0.6 around $u(46)$. Loosely stated, the stochastic effects

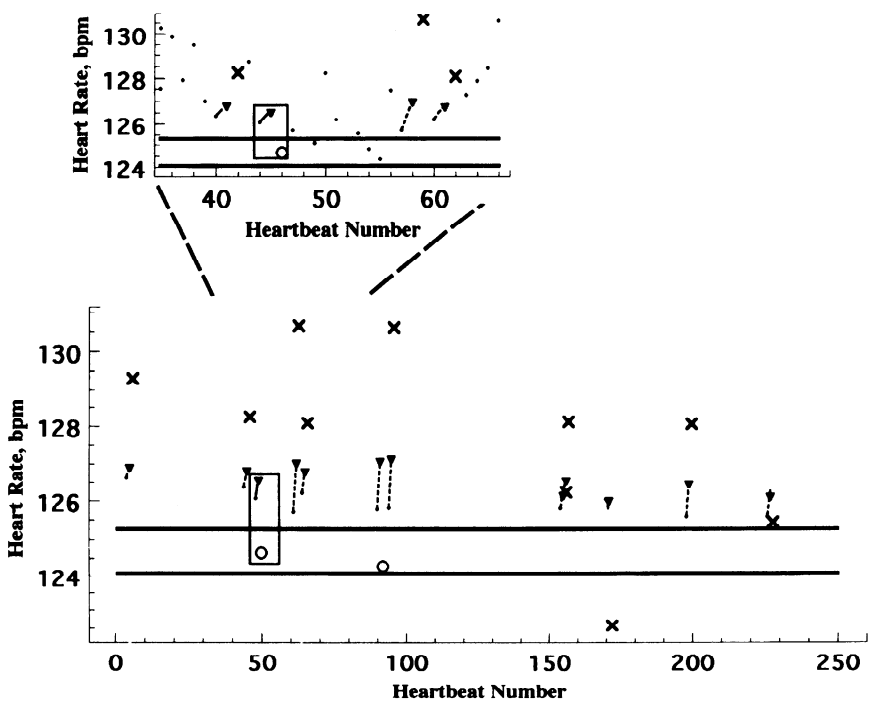

Fig. 4. Determination of those points $u(j + 2)$ that are close to $u(46)$, given that the conditioning vector $\mathbf{x}(j) = [u(j), u(j + 1)]$ is close to $\mathbf{x}(44) = [u(44), u(45)]$, for the normal infant HR tracing shown in Fig. 1*B*. Symbols and notation are identical to those used in Fig. 3. Excluding reference point $u(46)$, 1 candidate value of $u(j + 2)$ is close to $u(46)$, whereas 10 candidate values of $u(j + 2)$ are not close to $u(46)$. We thus form the conditional probability of the likelihood that runs of patterns that are close to the 2 observations commencing at $j = 44$ ($1 + 10 = 11$ conditioning vectors) remain close on next incremental comparison (1 successful point), as $1/11$. *Top*: close-up view near template vector $\mathbf{x}(44)$, including entire time series.

in this time series appear greater than those for the aborted-SIDS infant. This is exhibited in a close-up view (enclosed by a rectangle) of the template vector $\mathbf{x}(44)$ followed by $u(46)$. There has been a change in direction, in addition to a change in incremental length. Thus a candidate value of $u(j + 2)$ is in effect required to *1)* reverse direction from its predecessor, $u(j + 1)$, and then *2)* have incremental decrease that falls within the indicated band. This dual requirement is less likely to be met than the single requirement that the incremental change from $u(j + 1)$ to $u(j + 2)$ is approximately the same magnitude as the change from $u(45)$ to $u(46)$, if one could presume that directionality was not a factor.

Generally, in deterministic systems one would expect the change from the current to the next point to be approximately the same as the change from the immediately previous point to the current point and in the same direction. Stated in a formula, $u(j + 2) - u(j + 1)$ would approximately equal $u(j + 1) - u(j)$ with exact equality for linear systems (where ApEn = 0). The nonlinearity in some deterministic systems "causes" some differences in incremental changes and manifests itself visually in some candidate $u(j + 2)$ values missing the indicated band by undershooting or overshooting the predicted length change. Greater nonlinearity "causes" more undershooting or overshooting, smaller conditional probabilities, and thus greater ApEn. Greater stochastic influence "causes" more changes in direction, forcing the dual requirements indicated above to be satisfied more of the time, again producing smaller conditional probabilities and larger ApEn values. Both greater ensemble nonlinear and stochastic effects are manifested visually in greater randomness and complexity.

As an important technical note, observe that the tolerance $r = 0.6$ is the same in all component comparisons. Also, we do not care how close $u(44)$, $u(45)$, and $u(46)$ are to one another. Finally, the importance of not making r too small should now be evident, to ensure larger numbers of conditioning vectors and well-estimated conditional probabilities.

IMPLEMENTATION AND INTERPRETATION

The value of N for ApEn computations is typically between 100 and 5,000. Based on calculations that included both theoretical analysis (33, 38, 39) and clinical applications (21, 36, 37, 40), we have concluded that for $m = 2$ and $N = 1{,}000$, values of r from 0.1 to 0.25 SD of the $u(i)$ data produce good statistical validity of ApEn(m,r,N) for many models. For such r values, we demonstrated (33, 38, 39) the theoretical utility of ApEn(2,r,N) to distinguish data on the basis of regularity for both deterministic and random processes and the clinical utility in the aforementioned applications to HR data. These choices of m and r are made to ensure that the conditional probabilities defined in *Eq. 1* are reasonably estimated from the N input data points. Theoretical calculations indicate that reasonable estimates of these probabilities are achieved with an N value of at least 10^m and preferably at least 30^m points, analogous to a result for correlation dimension noted by Wolf et al. (46). For r values smaller than 0.1 SD, one usually achieves poor conditional probability estimates as well, whereas for r values larger than 0.25 SD, too much detailed system information is lost. In choosing $m = 2$ and $r = 15\%$ of the SD for the ApEn calculations for the data shown in Fig. 1 (Ref. 36), we conform to these guidelines.

For fixed m and r, the conditional probabilities given by *Eq. 1* are rigorously defined probabilistic quantities, marginal probabilities on a coarse partition, and contain a great deal of system information. Furthermore, these terms are finite and thus allow process discrimination for many classes of processes that have infinite Kolmogorov-Sinai (K-S) entropy (see RELATIONSHIP TO OTHER PARAMETERS). ApEn aggregates these (log) probabilities, thus requiring relatively modest data input. The key observation here is that fine reconstruction of the steady-state probability measures (attractors) is often not required to distinguish attractors from one another; comparison based on marginal probabilities will usually suffice to realize distinction.

ApEn is typically calculated via a short computer code. The form of ApEn provides for both de facto noise filtering, via choice of r, and artifact (outlier) insensitivity, via the probabilistic form of the comparisons. This robustness of ApEn to infrequent, even very large or small artifacts, in contrast to variability measures, is a useful statistical property for ambulatory applications.

Most importantly, despite the algorithm similarities, ApEn(m,r) is not intended to be an approximate value of K-S entropy (33, 37, 38). It is essential to consider ApEn(m,r,N) as a family of parameters; system comparisons are intended with fixed m and r. For a given system, there usually is significant variation in ApEn(m,r) over the range of m and r (36–38).

To avoid a significant contribution from noise in the ApEn calculation, one must choose r larger than most of the noise. Whereas exact guidelines depend on the distribution of the noise and the nature of the underlying system, we have had clinical success with r at least three times an estimated mean noise amplitude.

The physiological modeling of HR appears to be a very difficult problem; paradigms such as (complicated) correlated stochastic processes as well as mixtures of such processes and nonlinear deterministic systems are under active consideration as candidate models. The advantage of a broadly applicable parameter is that it can distinguish classes of systems for a wide variety of models. The mean, variability, and ApEn are all broadly applicable parameters in that they can distinguish many classes of systems, for which they can be meaningfully estimated from 1,000 data points. In applying ApEn, therefore, we are not testing for a particular model form, such as deterministic chaos; we are attempting to distinguish data sets on the basis of regularity (complexity). Such evolving regularity can be seen in both deterministic and random (stochastic) models (33, 38, 39).

To provide a potential physiological interpretation of what decreasing ApEn indicates, we propose a general hypothesis of perturbation and disease. A measure such as HR probably represents the output of multiple mecha-

nisms, including coupling (feedback) interactions such as sympathetic/parasympathetic response and external inputs ("noise") from internal and external sources. We hypothesize that a wide class of diseases and perturbations represent system decoupling and/or lessening of external inputs, in effect isolating a central system component from its ambient universe. ApEn provides a means of assessing this hypothesis, both for theoretical models and via actual data analysis. In general, ApEn increases with greater system coupling and greater external inputs, for a variety of models, thus providing an explicit index of autonomy in many coupled, complicated systems (33, 39). This observation supports our hypothesis, in conjunction with empirical results that have associated lowest ApEn values with aging or pathology (21, 36, 37, 40). The hypothesis also coincides with previous qualitative, graphic results, in that greater regularity (lower ApEn) generally corresponds to greater ensemble correlation in phase space diagrams and more total power concentrated in a narrow frequency range ("loss of spectral reserve"), both observed in a wide class of diseases (14, 15). The statistical utility of ApEn is highlighted in assessing this hypothesis, since complexity parameters such as the correlation dimension and K-S entropy cannot in general be used to differentiate competing models in which both stochastic and deterministic components evolve, as is indicated below.

RELATIONSHIP TO OTHER PARAMETERS

Variability

There is a fundamental difference between regularity parameters, such as ApEn, and variability measures: Most short- and long-term variability measures take raw HR data, preprocess the data, and then apply a calculation of SD (or of the similar, nonparametric percentile range variation) to the processed data (30). The means of preprocessing the raw data varies substantially with the different variability algorithms, giving rise to many distinct versions. Notably, once preprocessing the raw data is completed, the processed data are input for an algorithm for which the order of the data is immaterial. For ApEn, the order of the data is the crucial factor; discerning changes in order from apparently random to very regular is the primary focus of this parameter.

Variability measures, which assess the magnitude of deviation compared with a constant signal, have been shown to provide substantial insight in HR analysis; typically, significantly decreased variability has been correlated with adverse events (3, 22, 26, 42). In addition, there are settings in which variabilities are similar, whereas ApEn values are markedly different, which correspond to clinically consequential findings, illustrated in the comparison of HRs of the two infants made above. We thus see a utility in performing both variability and regularity analysis as part of an overall statistical protocol.

Parameters Related to Chaos

The historical development of mathematics to quantify regularity has centered around various types of entropy measures. Entropy is a concept that addresses system randomness and predictability, with greater entropy often associated with more randomness and less system order. Unfortunately, there are numerous entropy formulations, and many entropy definitions cannot be related to one other. K-S entropy, developed by Kolmogorov (24) and expanded on by Sinai, allows one to classify deterministic systems by rates of information generation. It is this form of entropy that algorithms such as those given by Grassberger and Procaccia (16) and by Eckmann and Ruelle (9) estimate. There has been keen interest in the development of these and related algorithms (e.g., Ref. 44) in the last ten years, since entropy has been shown to be a parameter that characterizes chaotic behavior (41).

However, we must recall that the K-S entropy was not developed for statistical applications and has major debits in this regard. For this reason, the K-S entropy is primarily applied by ergodic theorists to well-defined theoretical transformations, with no noise and an infinite amount of "data" available. K-S entropy is badly compromised by steady, small amounts of noise, generally requires a vast amount of input data to achieve convergence (29, 46), and is usually infinite for stochastic processes. These debits are key in the present context, since physiological time series, such as HR, are probably comprised of both stochastic and deterministic components.

ApEn was constructed along thematically similar lines to the K-S entropy, though with a different focus: to provide a widely applicable, statistically valid formula for the data analyst that will distinguish data sets by a measure of regularity (33, 37). The intuition motivating ApEn is that, if joint probability measures for reconstructed dynamics that describe each of two systems are different, then their marginal probability distributions on a fixed partition, given by conditional probabilities as in *Eq. 1,* are probably different. We typically need orders of magnitude fewer points to accurately estimate these marginal probabilities than to accurately reconstruct the "attractor" measure defining the process. ApEn has three technical advantages in comparison to K-S entropy for statistical usage. ApEn is nearly unaffected by noise of magnitude below r, the filter level; is robust to occasional, very large or small artifacts; gives meaningful information with a reasonable number of data points; and is finite for both stochastic and deterministic processes. This last point allows ApEn the capability to distinguish versions of stochastic processes from each other, whereas K-S entropy would be unable to do so.

There exists an extensive literature about understanding (chaotic) deterministic dynamical systems through reconstructed dynamics. Parameters such as correlation dimension (17), K-S entropy, and the Lyapunov spectrum have been much studied, as have techniques to utilize related algorithms in the presence of noise and limited data (7, 12, 27). Even more recently, prediction

(forecasting) techniques have been developed for chaotic systems (8, 10, 43). Most of these methods successfully employ embedding dimensions larger than $m = 2$, as is typically employed with ApEn. Thus in the deterministic dynamical system setting, for which these methods were developed, they are more powerful than ApEn in that they reconstruct the probability structure of the space with greater detail. However, in the general (stochastic, especially the "trickier" correlated stochastic process) setting, the statistical accuracy of the aforementioned parameters and methods appears to be poor, and the prediction techniques are no longer sensibly defined. Casdagli (8) is careful to apply his technique to a variety of test processes, most of which are noiseless deterministic systems, the others of which have small noise additively superimposed on a deterministic system. Complex stochastic processes (e.g., networks of queues, solutions to stochastic differential equations) are not evaluated, because they are not the systems under study in this literature. The relevant point here is that, because dynamical mechanisms of HR control remain undefined, a suitable statistic of regularity for HR must be more "cautious" to accommodate general classes of processes and their much more "diffuse" reconstructed dynamics.

We note that changes in ApEn generally agree with changes in dimension and entropy algorithms for low-dimensional, deterministic systems, consistent with expectation. The essential points here, assuring general utility, are that *1*) ApEn can potentially distinguish a wide variety of systems: low-dimensional deterministic systems, periodic and multiply periodic systems, high-dimensional chaotic systems, and stochastic and mixed (stochastic and deterministic) systems (33, 39) and *2*) ApEn is applicable to noisy, medium-sized data sets, such as those typically encountered in HR data analysis. Thus ApEn can be applied to settings for which the K-S entropy and correlation dimension are either undefined or infinite with good replicability properties as discussed below.

Statistical Validity: Error Bars for General Processes

The data analyst at this juncture can pose a critical question: how replicable are ApEn calculations? This is statistically addressed by SD calculations of ApEn, calculated for a variety of processes; such calculations provide "error bars" to quantify probability of true distinction. We (38, 39) performed calculations of the SD of ApEn(2, 15% process SD, 1,000) for two quite different template processes, the stochastic MIX(P) process defined in *The MIX Process: How Do We Tell Members Apart?* and the deterministic logistic map $f(x) = ax(1 - x)$, as a function of a. We determined, via Monte Carlo calculations, 100 replications per computation, that the SD of ApEn(2, 15% process SD, 1,000) < 0.055 for each P in MIX(P) and for each a in the logistic model. We also performed Monte Carlo calculations (100 replications/computation) for the first-order autoregressive AR(α,1) processes defined by $X(t) = \alpha X(t - 1) + Z(t)$, where we made a typical assumption that the $Z(t)$ are normally distributed independent, identically distributed (i.i.d.) random variables with variance $\sigma^2 = 1$. For each $\alpha > 0$, we determined that the SD of ApEn(2, 15% process SD, 1,000) was less than 0.045. As indicated in *Model Considerations and Error Bars*, we can expect that these SD bounds for ApEn will provide upper bounds for a large class of candidate models. It is this small SD of ApEn, applied to 1,000 points from various models, that provides its practical utility to HR data analysis. For instance, applying this analysis, we deduce that ApEn values that are 0.15 apart represent nearly 3 SDs distinction (assuming a Gaussian distribution of ApEn values, which seems to hold for most processes) (38), indicating true distinction with error probability of nearly $P = 0.001$. Similar statistical accuracy has not been established for dimension, K-S entropy, and Lyapunov exponent algorithms, in the general setting, as anticipated from the discussion in *Parameters Related to Chaos*.

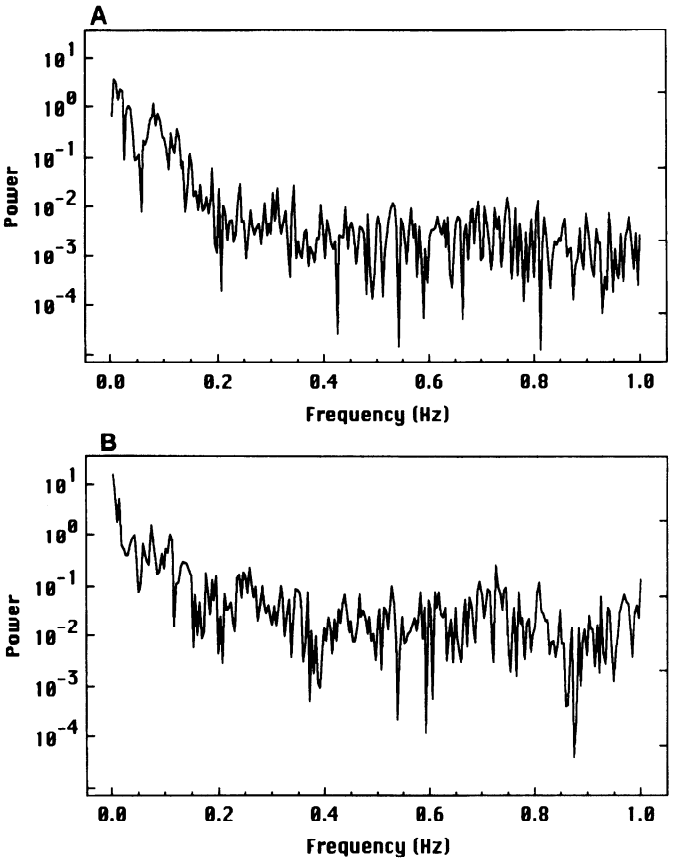

Fig. 5. Corresponding power spectra for HR data sets shown in Fig. 1 for an aborted-SIDS infant (*A*) and a normal infant (*B*).

Power Spectra

Examination of the power spectra corresponding to the two tracings in Fig. 1, shown in Fig. 5, also indicates a difference in the HR data, and provides a partial understanding of a general relationship between spectral information and ApEn. Both spectra are very broad banded and noisy for frequencies > 0.2 Hz. The primary difference between these spectra is that the tracing in Fig. 5*A* has more total power concentrated in a narrow frequency (here the low frequency 0–0.2 Hz) range,

whereas the tracing in Fig. 5*B* is more broad banded, with more power spread over a greater frequency range. Again, this interpretation implies greater regularity of *tracing A*: a spiked, narrowest-banded spectrum corresponds to a periodic function, e.g., a sinusoid. Of physiological note, the spectra do not appear to suggest that any particular frequency band (e.g., thermoregulatory, 0.05 Hz; baroreceptor feedback control, 0.1 Hz; respiratory, 0.1–0.3 Hz) evidences the greatest change between normal and aborted-SIDS infant data.

In general, smaller ApEn and greater regularity correspond to more ensemble dependence in time-series and process characterization. The two opposing extremes are *1*) periodic and linear deterministic models, which produce very peaked, narrow-banded spectra with low ApEn values and *2*) sequences of independent random variables, for which time series yield intuitively highly erratic behavior and for which spectra are very broad banded with high ApEn values. Intermediate to these extremes are deterministic, nonlinear chaotic models and correlated stochastic processes, both of which can exhibit complicated spectral behavior. In some instances, the comparison in the spectral domain of healthy and diseased states may be crucial when pronounced differences in a particular frequency band suggests an underlying specific physiological disorder. In other instances, such as in the comparison of the two infants in Fig. 5, there is more of an ensemble difference between the time series, viewed in both the time domain and the frequency domain, and the need remains to encapsulate the ensemble information into a single value, a replicable test statistic, to distinguish the data sets.

The MIX Process: How Do We Tell Members Apart?

Calibrating statistical analysis to intuitive sensibility for the MIX(P) process (33) highlights many important points discussed above. MIX is a family of stochastic processes that samples a sine wave for $P = 0$, consists of i.i.d. samples selected uniformly ("completely randomly") from an interval for $P = 1$, intuitively becoming more random as P increases, as illustrated in Fig. 6. Formally, to define MIX(P), first fix $0 \leq P \leq 1$. Define $X_j = \sqrt{2}\sin(2\pi j/12)$ for all j, Y_j = i.i.d. uniform random variables on $[-\sqrt{3}, \sqrt{3}]$, and Z_j = i.i.d. random variables, $Z_j = 1$ with probability P, $Z_j = 0$ with probability $1 - P$. Then define $\text{MIX}(P)_j = (1 - Z_j)X_j + Z_jY_j$. A mechanistic motivation for the MIX(P) process is indicated in a previous work (33); the appellation MIX indicates the formulation of this process as a composite, or mixture, of deterministic and stochastic components.

As P increases, the process becomes apparently more irregular, unpredictable, or complex, and we would like a statistic that quantifies this evolution. The MIX process has mean 0 and SD 1 for all P, so these moments do not discriminate members of MIX from one another. Furthermore, the correlation dimension of MIX(P) = 0 for $P < 1$, and the correlation dimension of MIX(1) = ∞ (33). In addition, the K-S entropy of MIX(P) = ∞ for $P > 0$, and = 0 for MIX(0). Thus both of these chaos statistics perform terribly for this template process,

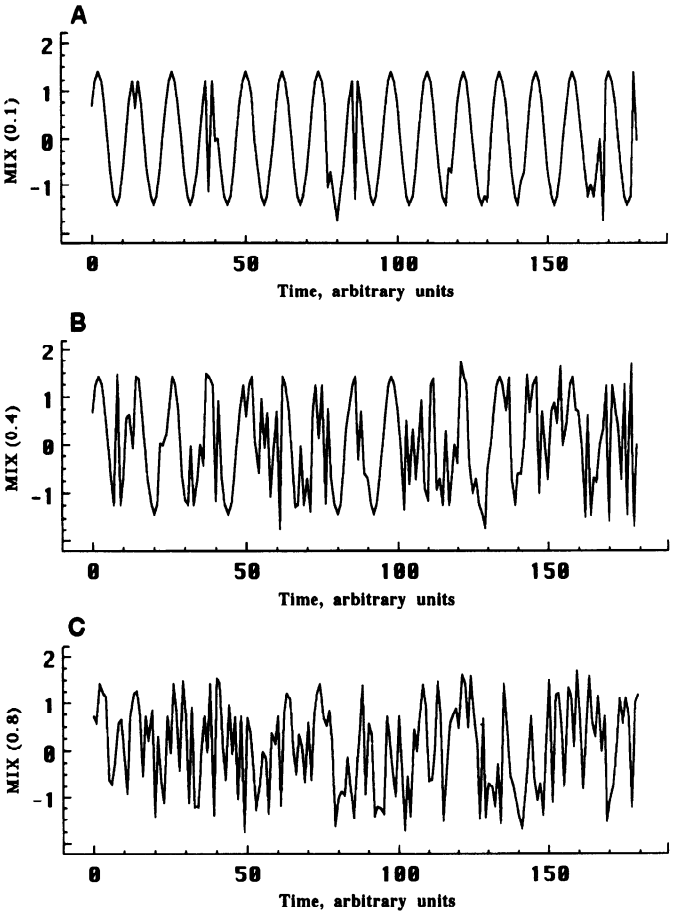

Fig. 6. MIX(P) model time-series output for 3 parameter values: P = 0.1 (*A*), P = 0.4 (*B*), and P = 0.8 (C). MIX(P) is a family of processes that samples a sine wave for $P = 0$, samples independent, identically distributed (i.i.d.) uniform random variables for $P = 1$, and intuitively becomes more irregular as P increases. ApEn quantifies the increasing irregularity and complexity with increasing P: for typical parameter values N = 1,000 points, m = 2, and r = 0.18, ApEn [MIX(0.1)] = 0.436, ApEn [MIX(0.4)] = 1.455, and ApEn [MIX(0.8)] = 1.801. In contrast, correlation dimension of MIX(P) = 0 for all $P < 1$, and the K-S entropy of MIX(P) = ∞ for all $P > 0$. Thus even given no noise and an infinite amount of data, these latter two measures do not discriminate the MIX(P) family.

even given an infinite amount of theoretical data: they do not distinguish members of MIX(P) from one another, do not capture the evolving complexity change, and are discontinous with the parameter P (near $P = 1$ for the correlation dimension, near $P = 0$ for K-S entropy). We anticipate that many of the difficulties of the correlation dimension and the K-S entropy applied to MIX(P) would be mirrored for other correlated stochastic processes. Stated differently, even if an infinite number of points were available, with no added "noise" atop MIX(P), we still could not use the correlation dimension or K-S entropy to distinguish members of this family. The difficulty here is that the parameters are identical, not that we have insufficient data or too much noise. In contrast to K-S entropy, and in conjunction with intuition, ApEn increases with increasing P, as indicated in the legend to Fig. 6.

From another viewpoint, for $P < 1$ we have that the correlation dimension of MIX(P) is finite (= 0), yet

MIX(P) is a stochastic process. Thus a finite value of correlation dimension does not necessarily indicate determinism. Also, the finiteness of ApEn for MIX(P), in contrast to the infinite value of the K-S entropy for this family, reinforces the point that ApEn is not intended as an estimate of the K-S entropy but rather as an autonomous quantity.

Finally, evaluation of MIX(P) raises another point: an effort to separate signal from "noise" is often addressing the wrong question in data analysis of mixed (composite) systems, where both deterministic and stochastic components could be varying, as in MIX. What is evolving here is the relative weighting of the stochastic to the deterministic component. Separation of signal and noise here would indicate a true "signal" of a sine wave, with an overlay of i.i.d. uniform noise for all P, but it would not be until some assessment was made of how much noise was present relative to the base that we could distinguish members of this family. In effect, ApEn accomplishes this.

CHOICE OF INPUT PARAMETERS AND STATISTICAL AND MODEL-RELATED ISSUES

We now discuss further mathematical and statistical properties of ApEn. This section is at times necessarily more technical than the remainder of this article.

Parameter Considerations

Informational-theoretical generalization. The rationale for ApEn is to provide information-theoretical parameters for continuous-state processes. For probability measures, the differential entropy provides such a parameter; for comparing two measures, the Kullback-Leibler information serves this purpose. For processes, however, the story is much more limited. This notion is quantified by the rate of entropy in the discrete state space, Markov chain case (Theorem 3.2.8 in Ref. 5), and by the aforementioned K-S entropy for deterministic differential equations and dynamical systems. Notably, for both of these process definitions, the straightforward extension to general processes yields a value of infinity for most stochastic, continuous state-space processes, thus rendering such extensions useless as a means of distinguishing many processes of interest.

The main analytic property of ApEn is that ApEn(m,r) is the continuous state-space extension of the rate of entropy, in that one recovers the rate of entropy in the discrete state, Markov chain settings, provided the state-space elements are at least distance r from each other. This result is established both for general Markov chains (Theorem 3 in Ref. 33) and for approximating (m,r) Markov chains to a given process (Theorem 1 in Ref. 35). Specifically, in these settings $\mathrm{ApEn}(m,r) = -\Sigma_{x\in X}\ \Sigma_{y\in X}\ \pi(x)p_{xy}\log(p_{xy})$, where $\pi(x)$ is the steady-state probability of state $\{x\}$ and p_{xy} is the transition probability from state $\{x\}$ to state $\{y\}$.

The orientation of ApEn is to discriminate processes via a particular notion (regularity) rather than to simply discriminate them. If one wishes to discriminate two processes, given the conditional (cell transition) probabilities that an mth order embedding gives on cells of width r, then there are two rather different, successful approaches that have well-understood properties in appropriate settings. The methods of Billingsley (4) as well as collaborators (Darwin, Anderson, and Goodman) to discriminate known Markov chains with s states yields a χ^2 statistic with $s(s-1)$ degrees of freedom for testing if two Markov chains are distinct. Alternatively, in information theory, classification algorithms that are based on universal data compression schemes, such as the universal discriminant function given by Ziv (Eq. 23 in Ref. 47), have been seen to be effective (exponentially optimal as the sample length tends to infinity) for general finite state-space processes with a small alphabet.

Mesh interplay. How we choose ApEn input parameters m and r raises not only statistical estimation issues, but also first and more fundamentally, parameter issues. In brief, the interplay between meshes need not be nice, in general, in ascertaining which of (two) processes is "more" random, discussed below. Also, we need to determine whether observed noise is (physiologically) real or measurement error to be ignored. Compare, for example, *process A* [MIX(0.1) superimposed by i.i.d. noise, uniformly distributed in the range $(-0.01, 0.01)$] and *process B* [MIX(0.4) superimposed by i.i.d. noise, uniformly distributed in the range $(-0.01, 0.01)$]. Visual inspection tacitly filters out the low-magnitude i.i.d. noise contributions, and thus one might conclude that *process A* is more regular than *process B*. This viewpoint is consistent with what ApEn calculations show, provided that r is somewhat larger than 0.01, the noise level. In contrast, for any $r < 0.001$, *processes A* and *B* have nearly identical ApEn(m,r) values: we are primarily seeing the noise. Both of these calculations calibrate with intuition in that, if the noise is to be ignored as nonphysiological, then we should choose $r > 0.01$, whereas if the noise is physiologically significant, we should choose r smaller than the typical noise level.

In general, we might like to ask the parameter question: given no noise and an infinite amount of data, can we say that *process A* is more regular than *process B*? First, note that for many (e.g., i.i.d., totally random) processes, ApEn(m,r) grows with decreasing r like $\log(2r)$ (38). Thus for many processes, as $r \to 0$, ApEn(m,r) diverges to ∞; thus we cannot answer the parameter question by comparing limiting ApEn(m,r)(A) and ApEn(m,r)(B) values as $r \to 0$. The "flip-flop pair" of processes (38) implies that the answer to the parameter question is "not necessarily": in general, comparison of relative process randomness at a prescribed level is the best that one can do. That is, processes may appear more random than processes on many choices of partitions, but not necessarily on all partitions of suitably small diameter. The flip-flop pair are two i.i.d. *processes A* and *B* with the property that for any integer m and any positive r, there exists $s < r$ such that ApEn(m,s)(A) < ApEn(m,s)(B) and there exists $t < s$ such that ApEn(m,t)(B) < ApEn(m,t)(A). At alternatingly small levels of refinement given by r, *process B* appears more random and less regular than *process A* followed by

appearing less random and more regular than *process A* on a still smaller mesh (smaller r). In this construction, r can be made arbitrarily small, thus establishing the point, that process regularity is a relative [to mesh, or (m,r) choice] notion.

Relative consistency. For many *processes A* and *B*, we can assert more than relative regularity, even though both *processes A* and *B* will typically have infinite K-S entropy. For such pairs of processes, which have been denoted as a completely consistent pair (38), whenever $\mathrm{ApEn}(m,r)(A) < \mathrm{ApEn}(m,r)(B)$ for any specific choice of m and r, then it follows that $\mathrm{ApEn}(n,s)(A) < \mathrm{ApEn}(n,s)(B)$ for all choices of n and s. Any two elements of {MIX(*P*)}, for example, appear to be completely consistent. The importance of completely consistent pairs is that we can then assert that *process B* is more irregular (or random) than *process A* without needing to indicate m and r. Visually, *process B* appears more random than *process A* at any level of view. We anticipate that the following conjecture is relatively straightforward to prove, giving a sufficient condition to ensure that *processes A* and *B* are a completely consistent pair and indicating the relationship to the autocorrelation function.

Conjecture: let A and B be continuous-time, stationary stochastic processes with autocorrelation functions $acf(t)$ and $acg(t)$, respectively, such that *1)* $acf(t) > acg(t) > 0$ for all $t > 0$ and *2)* the variance of A = variance of B. Pick any time-sampling increment $\Delta t > 0$, and form associated discrete-time stationary *processes* $A(\Delta t)$ and $B(\Delta t)$ from A and B. We conclude that for any Δt, $A(\Delta t)$ and $B(\Delta t)$ are a completely consistent pair, with $B(\Delta t)$ appearing more random (higher ApEn values) than $A(\Delta t)$ for all m and r.

We reemphasize that the utility of ApEn is as a relative measure, with fixed m and r. We see no sensible comparisons of $\mathrm{ApEn}(m,r)(A)$ and $\mathrm{ApEn}(n,s)(B)$ for *processes A* and *B* unless $m = n$ and $r = s$. Both theoretically (e.g., logistic map, Henon map from dynamical systems) and on observed data, we often observe a relative consistency of ApEn over a statistically valid range of (m,r) pairs, similar to that given by completely consistent pairs; whenever the statistical estimate $\mathrm{ApEn}(m,r,N)(A) < \mathrm{ApEn}(m,r,N)(B)$ for an (m,r) pair, then $\mathrm{ApEn}(n,s,N)(A) < \mathrm{ApEn}(n,s,N)(B)$ for all (n,s) pairs in the range.[4] Clinically, we determined that the association of very low (HR) ApEn values with aborted-SIDS infants was replicated for different choices of parameter values m and r for ApEn input, even though the ApEn values themselves changed markedly with different m and r choices (36). This relative consistency property was also confirmed via analysis of 15 (m,r) pairs in a study of healthy and severely ill neonates (37).

These findings thus impart a robustness to ApEn input parameter choice insofar as the identification of time series with atypical ApEn values.

Analytic expressions. For many processes, we can provide analytic expressions for $\mathrm{ApEn}(m,r)$. Two such expressions are given by the two following theorems (see Ref. 35). *Theorem 1:* assume a stationary process $u(i)$ with continuous state space. Let $\mu(x, y)$ be the joint stationary probability measure on R^2 for this process, and $\pi(x)$ be the equilibrium probability of x. Then $\mathrm{ApEn}(1,r) = -\int \mu(x,y) \log [\int_{z=y-r}^{y+r} \int_{w=x-r}^{x+r} \mu(w,z)\, dw\, dz / \int_{w=x-r}^{x+r} \pi(w)\, dw]\, dx\, dy$. *Theorem 2:* for an i.i.d. process with density function $\pi(x)$ (for any $m \geq 1$), $\mathrm{ApEn}(m,r) = -\int \pi(y) \log [\int_{z=y-r}^{y+r} \pi(z)\, dz]\, dy$.

Theorem 1 can be extended in straightforward fashion to derive an expression for $\mathrm{ApEn}(m,r)$ in terms of the joint [$(m + 1)$-fold] probability distributions. Hence we can calculate $\mathrm{ApEn}(m,r)$ for Gaussian processes, since we know the joint probability distribution in terms of the covariance matrix. This important class of processes (for which finite sums of discretely sample variables have multivariate normal distributions) describes many stochastic models, including solutions to autoregressive-moving average (ARMA) models, and linear stochastic differential equations driven by white noise.

Thus for specified m and r, $\mathrm{ApEn}(m,r)$ is a well-defined process parameter, just as the K-S entropy is. ApEn does not have two properties that the K-S entropy has: *1)* K-S entropy establishes process isomorphism (Ornstein's theorem) (28) and *2)* a positive K-S entropy value establishes chaos when applied to deterministic, dynamical systems (9), but neither of these is a question motivated by statistical application, and the K-S entropy has the statistical debits indicated above in *Parameters Related to Chaos.*

Statistical Considerations

Choice of input parameters: curse of dimensionality. To understand the statistical requirements imposed by ApEn, we can think of an (m,r) choice of input parameters as partitioning the state space into uniform width boxes of width r, from which we estimate mth-order conditional probabilities. [This is exactly the case if the state space is discrete, with evenly-spaced state values, by *Theorem 1* (35). Otherwise, this description approximates what ApEn does; ApEn is defined in a less ad hoc fashion in that the partition would be "floated" according to the stationary measure rather than fixed, thus avoiding some partition boundary arbitrariness]. For state space $[-A/2, A/2]$, we would have $(A/r)^{m+1}$ conditional probabilities to estimate. Specifically, divide $[-A/2, A/2]$ into A/r cells; the ith cell $\mathrm{C}(i) = [x, x + r)$, where $x = -(A/2) + (i - 1)r$. Then define the conditional probability $p_{\mathrm{ivect},j}$ for all length m vectors of integers ivect and integers j, ivect $= (i_1,i_2,\ldots, i_m)$, $1 \leq i_k \leq A/r$ for all k, $1 \leq j \leq A/r$, by $p_{\mathrm{ivect},j}$ = {conditional probability that $u(k) \in \mathrm{C}(j)$, given that $u(k-1) \in \mathrm{C}(i_1)$, $u(k-2) \in \mathrm{C}(i_2)$, ..., and $u(k-m) \in \mathrm{C}(i_m)$}. We assume stationarity in defining these conditional probabilities; in the very

[4] Thematic of the conjecture, we expect that such statistical consistency will correspond to the case in which the autocorrelation functions $acf(t)$ and $acg(t)$ satisfy $|acf(t)| > |acg(t)|$ for all t in some range $t_1 < t < t_2$. Also, in the above conjecture, we can accommodate the setting in which *processes A* and *B* have different variance by proving the result for normalized ApEn (21, 36), in which r will be a fixed percentage of the respective process variances.

general ergodic case, these conditional probabilities are given by limits of time averages.

For general stochastic processes, many of these conditional probabilities will be nonzero, so we need to accommodate reasonable estimates of the $(A/r)^{m+1}$ conditional probabilities given N data points. If m is relatively large, or if r is too small, the number of probabilities to be estimated will be unwieldy, and statistical estimates will be poor for typical data set lengths. We are thus required to have the mesh (given by r) coarse enough, and the embedding dimension (given by m) low enough to handle the limits of resolution of the data, to ensure replicability of conditional probability estimates.

A choice of $m = 2$ is superior to $m = 1$, in that it allows more detailed reconstruction of the joint probabilistic dynamics of the process. We typically cannot use $m > 2$, due to a combination of two factors: *1*) to ensure homogeneity of the subject's state, the number of input points (e.g., heart beats) N often cannot exceed 5,000, and *2*) once the constraint on N is set, $m > 2$ produces poor conditional (stationary) probability estimates unless r is very large, and such large r values generally are too coarse to realize pronounced process distinctions via ApEn(m,r).

If two *processes* A and B agree on relatively coarse meshes with small embedding dimension m, we would require large N to reasonably estimate the mth order conditional probabilities on a fine mesh (or of a high order) to potentially distinguish *processes* A and B. This observation is analogous to the result for polynomials. Given Taylor expansions for analytic functions f and g, if they differ in the linear or quadratic term significantly, then we expect to distinguish f and g with few values; if the polynomials agree, e.g., to eighth order, then we generally require many more points, assuming a small amount of measurement uncertainty.

The extent to which rapidly growing data lengths $N(m)$ are required to realize accurate parameter estimation given embedding dimension m is highlighted in the related problem of approximating a Bernoulli process (both in Ornstein's "d-bar" metric and for the K-S entropy). Ornstein and Weiss indicate a "guessing scheme" (29) to perform this approximation to arbitrary accuracy: they require a supergeometric number of points $N(m)$ as a function of the embedding dimension to ensure convergence of the method. Precisely, they require a technical improvement of the Shannon-Breiman-McMillan theorem (6) for which $\lim_{m\to\infty} N(m)/A^m = \infty$ for all $A > 0$.

Bias. ApEn is a biased statistic (although we anticipate asymptotically unbiased for many processes); the expected value of ApEn(m,r,N) increases asymptotically with N to ApEn(m,r), for all processes that we have studied. Thus to ensure appropriate comparisons between data sets, N must be the same for each data set. This bias arises from two separate considerations. The first source is the concavity (and nonlinearity) of the logarithm function in the ApEn definition. In general, unbiased estimators are uncommon in nonlinear estimation; this observation also applies to algorithms estimating the correlation dimension and K-S entropy for dynamical systems. More significant is the second source of the bias. In the definition of $C_i^m(r)$ in forming ApEn, the template vector $\mathbf{x}(i)$ itself counts in the $C_i^m(r)$ aggregation of vectors close to $\mathbf{x}(i)$. This is done to ensure that calculations involving logarithms remain finite, but has the consequence that the conditional probabilities estimated in *Eq. 1* are underestimated. Operationally, the inclusion of the template vector in the $C_i^m(r)$ count adds 1 to both numerator and denominator counts as performed in QUANTIFICATION OF REGULARITY (and algorithmically) in all conditional probability estimates. If there are few $\mathbf{x}(j)$ within r of $\mathbf{x}(i)$ for many template vectors $\mathbf{x}(i)$, this procedure can result in a bias of 20–30% in the ApEn statistic (e.g., see Fig. 1 in Ref. 38).

The estimator $\mathrm{ApEn}(m,r,N) = \Phi^m(r) - \Phi^{m+1}(r)$ was chosen to agree with historical precedent; the algorithms for correlation dimension (17) and the K-S entropy (Eq. 5.7 in Ref. 9) also compare the template vector to itself in ensemble comparisons. For fixed m and r, the effect of this component of the bias tends to 0 as $N \to \infty$, but for typical length data sets (e.g., N = 1,000), the bias must be considered. It should thus be noted that the SD calculations for ApEn given above in *Statistical Validity: Error bars for General Processes* were for the ApEn(2,r,1000) statistic; for larger N, we would expect smaller ApEn SD, whereas for smaller N, we would expect larger ApEn SD. Importantly, the difference E[ApEn(2,r,1000)] − ApEn(2,r), where E indicates expectation, is significantly larger than the SD of ApEn(2,r,1000), so estimation of the parameter ApEn(2,r) based on the mean observed value of ApEn(2,r,1000) would be inappropriate without a correction term for the bias. We propose that the following ϵ estimators $\mathrm{ApEn}_\epsilon(m,r,N)$ may be effective in bias reduction for ApEn: fix ϵ, and define $C_{i,\epsilon}^m(r) = (\epsilon$ + no. of j unequal to $i \le N - m + 1$ such that $d[\mathbf{x}(i),\mathbf{x}(j)] \le r)/(N - m + 1)$.

Then define $\Phi_\epsilon^m(r)$ as $(N - m + 1)^{-1} \sum_{i=1}^{N-m+1} \ln C_{i,\epsilon}^m(r)$ and $\mathrm{ApEn}_\epsilon(m,r,N)$ as $\Phi_\epsilon^m(r) - \Phi_\epsilon^{m+1}(r)$.

We thus markedly diminish the effect of comparing the test vector to itself by giving weight ϵ to this comparison. Procedures for choosing ϵ would depend on N; asymptotic evaluations of the bias reduction would require some knowledge of the u process. Similar ϵ estimation could also be applied to the aforementioned parameters related to chaos.

Operational consequences of bias. We can now operationally see the consequences of applying "chaos" algorithms to general stochastic processes. We compare two processes, MIX(0.5) and logis(3.6), the logistic map $f(x) = 3.6x(1 - x)$, considering an embedding dimension of $m = 3$, and a moderately coarse mesh width r that subdivides the state spaces into 20 cells of equal width [thus $r = \sqrt{3}/10$ for MIX(0.5), whereas $r = 0.05$ logis(3.6)]. Both MIX(0.5) and logis(3.6) are (auto)correlated processes, and we want to estimate the conditional probabilities $p_{\mathrm{ivect},j}$ for all length 3 vectors ivect and integers $1 \le j \le 20$ for each process. For logis(3.6), only 38 of the 20^4 conditional probabilities are nonzero, and

thus each of these can be reasonably estimated given moderate-sized data sets. For instance, given $N = 1{,}000$, mimicking the procedure indicated in section 2, we find that there are 41 instances (values of k) in which triples of contiguous points $(u(k-1), u(k-2), u(k-3))$ satisfy $\{u(k-1) \in C(18),\ u(k-2) \in C(12)$, and $u(k-3) \in C(16)\}$. Of these 41 conditioning vectors, 30 4-tuples of points $(u(k), u(k-1), u(k-2), u(k-3))$ satisfy $\{u(k) \in C(9), u(k-1) \in C(18), u(k-2) \in C(12)$, and $u(k-3) \in C(16)\}$; we then estimate the conditional probability $p_{(18,12,16),9} = {}^{30}/_{41}$. A second estimate based on a subsequent 1,000 point sequence would probably produce a similar estimate for $p_{(18,12,16),9}$, and adding 1 to both numerator and denominator in these estimates (e.g., to form ${}^{31}/_{42}$ as the conditional probability estimate) would not have a pronounced effect. It is precisely the (increasing and extreme) sparseness of the set of $(m+1)$-tuples for which the true conditional probabilities $p_{\text{ivect},j}$ are nonzero for true dynamical systems, as a function of embedding dimension m and mesh width (scaling range) r, that affords large m and small r in conditional probability estimation (reconstruction) for these processes.

In contrast, for MIX(0.5), all 20^4 (= 160,000) conditional probabilities $p_{\text{ivect},j}$ are nonzero, and require estimation. For most (ivect, j) pairs, the probabilities will be rare events, nearly 0; for $N = 1{,}000$ points, the number of conditioning vectors that match ivect will typically be 0, 1, 2, or 3, for most ivect 3-tuples. As indicated in *Bias*, comparison of the template vector to itself would then operationally add 1 to both numerator and denominator, yielding conditional probability calculations of $(0+1)/(0+1)$, $(0+1)/(1+1)$, $(0+1)/(2+1)$, or $(0+1)/(3+1)$ (thus of 1, ½, ⅓, or ¼) as extremely poor estimates of the true very small conditional probabilities. This effect is heightened as m grows larger and as r shrinks and thus operationally clarifies the point made above, that for general stochastic processes, there is poor replicability (estimates of many conditional probabilities, and hence) of the probability structure given moderate sized N, for either high embedding dimension m or small scaling range r.

Similarly, we see the operational effects of nonstationarity, especially trending, on the ApEn(m,r,N) statistic (and on other complexity algorithms). Trending will typically produce a reduction in the number of conditioning vectors close to many template vectors, thus enhancing the effect of comparing the template vector to itself in the $C_i^m(r)$ definition. Again, this will spuriously lower the ApEn(m,r,N) estimate.

Asymptotic normality. ApEn(m,r,N) appears to exhibit asymptotically normal distributions for general (weakly dependent) processes (38). However, analytic proofs of asymptotic normality, and especially explicit variance estimates for ApEn(m,r,N), appear to be extremely difficult. In theory, one wishes to establish appropriate central limit theorems (CLTs), from which normality would follow. However, the terms $C_i^m(r)$ that form the basic components of the ApEn definition are complicated random variables with contributions from all the random variables $u(i)$. Hence, for even the simplest processes that define the $\{u_i\}$, CLTs will have to accommodate a high degree of correlation among the summands. Both Peligrad's CLT (31) and analogous results for U statistics (25) are candidate tools for the establishment of CLTs in this setting; these approaches require suitable mixing hypotheses, with Peligrad's conditions more specific and her results more explicit.

Model Considerations and Error Bars

Insufficiency of existing general network models. In modeling general classes of real biologic networks, we anticipate that any single model form, such as the MIX(P) process, deterministic differential equations, or ARMA models is inadequate. At the least, we would expect faithful models in many settings to incorporate queueing network and (adaptive) control theory considerations, as well as the likelihood that different nodes in a network are defined by distinct mathematical models. Queueing models arise naturally in multinode network analysis with interconnections; control models arise from considering the brain (or some focal component) as an intelligent central processor, possibly altering system characteristics based on, e.g., a threshold response. Queueing theory has developed largely within communication (traffic) theory (18) and computer network analysis (2, 19), whereas control theory has developed toward optimizing performance in engineering systems (11). Notably, analytic developments from these fields may not be directly suitable to physiological network modeling, not surprisingly, since these fields were not driven by biologic context. Two physiologically-motivated problems within these fields that seem worthy of significant effort are to describe the behavior of *1*) queueing networks in which some nodes are (coupled) deterministic oscillators and *2*) (adaptive) control systems in which there is a balking probability P with which the control strategy is not implemented. The second of these problems could model some diseases in which messages may not reach the controller or the controller may be too overwhelmed to respond as indicated.

Limitations of error bars. Although we recognize that it is a very difficult problem to accurately model, e.g., HR, we still need to have some sense of the error in ApEn estimation to ensure reproducibility. The models for which we calculated error bars [logistic map, MIX(P), AR(α,1)] were chosen as representative of low-order and "weak"-dependence processes, which characterize most familiar analytic models. Furthermore, the MIX(P) and AR(α,1) computations are appealing, in that the MIX(P) process is nearly i.i.d. for P near 1, and the AR(α,1) process is nearly i.i.d. for α near 0. Because larger ApEn SD often corresponds to more uncorrelated processes, we expect that the SD bounds for ApEn for MIX(P) and for AR(α,1) will provide upper bounds for a large class of (weakly dependent) processes.

For "strong"-dependence processes, one often requires huge numbers of samples to estimate any parameter (e.g., mean, SD, ApEn) with good statistical precision. Informally, one does not see a representative piece of the steady-state measure without huge sample size N. Queues in heavy traffic are prototypal strong-depen-

dence processes, which can be seen from the form of heavy-traffic limit theorems; Whitt (Ref. 45, especially Eq. 53) provides a formula for approximate sample size required for typical statistical accuracy in open queueing networks, and in very heavy traffic, this sample size can be $N > 1{,}000{,}000$. Thus in instances in which queueing models in heavy traffic accurately model (a component of) a physiological network, similarly huge numbers of data points will be required.

CAVEATS AND CONCLUSIONS

To ensure appropriate ApEn application and interpretation, we conclude with the following caveats.

1) If system noise is very large, e.g., consistent signal-to-noise ratio < 3, the validity of ApEn and many other statistical calculations may be seriously compromised. Also, for specified m (window length) in the ApEn calculation, it is important to have at least 10^m, preferably 20^m data points to analyze. Otherwise the conditional probability estimates that form the "backbone" of the ApEn calculation may not be accurate.

2) ApEn is a regularity, not a magnitude statistic. Whereas it affords a new approach to data analysis, it does not replace moment statistics, such as the mean and SD. As such, we recommend use of ApEn in conjunction with other statistics, not as a sole indicator of system characteristics.

3) ApEn decrease often correlates with SD decrease. This is not a "problem," as statistics often correlate with one another, but there are times when we desire an index of regularity decorrelated from SD. We can realize such an index, by specifying r in ApEn(m,r,N) as a fixed percentage of the sample SD of the individual (not group, or fixed) subject data set, with possibly a different r for each subject (21, 36). We call this normalized regularity. Its utility was demonstrated, e.g., in refining the differences between HR data from quiet and rapid-eye-movement (REM) infant sleep, two sleep states with markedly different overall variability (SD), by establishing that REM sleep has significantly greater normalized regularity (smaller normalized ApEn values) than quiet sleep has (36), which is juxtaposed with the more classical finding of much larger variability in REM sleep than in quiet sleep. This thus enables us to determine whether two processes are different both in variation (SD) and in normalized regularity.

4) We anticipate that the primary utility of ApEn will be to uncover subtle abnormalities or alterations in long-term data that are not otherwise apparent. We do not anticipate that it will be so useful in discerning acute changes that are associated with relatively infrequent data measurements. This again follows from the probabilistic nature of *Eq. 1*, which implies that no single, very small subset of the data can dramatically affect ApEn (unlike moment statistics).

5) If the (HR) time series is nonstationary, that is, contains one or more pronounced trends upward or downward, little can be inferred from moment (mean, variability), ApEn, or power spectral calculations, because the trends tend to dominate all other features. Physiologically, data with trends suggest a collection of heterogeneous epochs as opposed to a single homogeneous state. From the statistical perspective, it is imperative that any trends be removed before meaningful interpretation can be made from statistical calculations.

6) Whereas chaotic models for time series are often appropriate, especially when the underlying system behavior is derived from (a set of) differential equations, we can rarely, if ever, infer chaos or determinism solely from data considerations. Proper interpretation of ApEn is as a regularity statistic to distinguish finite, noisy, possibly stochastic or composite deterministic and stochastic data sets, not to infer model form.[5]

This work was supported in part by National Heart, Lung, and Blood Institute Grant HL-42172, National Aeronautics and Space Administration Grant NAG2–514, the Mathers Charitable Foundation, and National Institute on Drug Abuse Grant DA-06306.

Address for reprint requests: S. M. Pincus, 990 Moose Hill Rd., Guilford, CT 06437.

Received 30 December 1992; accepted in final form 29 October 1993.

[5] Several myths exist vis-à-vis data and parameter interpretation that must be corrected. First, correlation in phase space does not imply determinism (consider the MIX process or autoregressive or moving average models extensively studied in statistics). Second, correlation dimension does not indicate the number of free parameters needed to model a system. A sequence of independent random variables, all taken from Gaussian random variables, mean μ, variance 1 will produce a correlation dimension of ∞ for all μ, yet this system is specified by one free variable, the value of μ. The desire for a low-parameter model of system behavior is apparent, to specify a system via a minimum number of inputs, but interpretation of correlation dimension in terms of numbers of degrees of freedom for stochastic models or their resultant data sets seems a dubious proposition. Furthermore, differentiation of the aforementioned Gaussian with $\mu = 0$ from that with $\mu = 10$ is established with very high probability from 1,000 data points, a classical statistical result, so correlation dimension cannot be utilized to determine the number of data points needed to distinguish competing stochastic models. Finally, a strange attractor or fractal steady-state structure does not necessarily imply a deterministic model (e.g., see Ref. 32 for fractal behavior arising from stochastic random matrices and Theorem 6 in Ref. 35, which mathematically proves that any steady-state measure arising from a deterministic dynamical system model can be approximated to arbitrary accuracy by that from a stochastic Markov chain).

REFERENCES

1. **Aarimaa, T., R. Oja, K. Antila, and I. Valimaki.** Interaction of heart rate and respiration in newborn babies. *Pediatr. Res.* 24: 745–750, 1988.
2. **Allen, A. O.** *Probability, Statistics, and Queueing Theory: With Computer Science Applications* (2nd ed.). San Diego, CA: Academic, 1990.
3. **Bigger, J. T., Jr., M. T. LaRovere, R. C. Steinman, J. L. Fleiss, J. N. Rottman, L. M. Rolnitzky, and P. J. Schwartz.** Comparison of baroreflex sensitivity and heart period variability after myocardial infarction. *J. Am. Coll. Cardiol.* 14: 1511–1518, 1989.
4. **Billingsley, P.** Statistical methods in Markov chains. *Ann. Math. Stat.* 32: 12–40, 1961.
5. **Blahut, R. E.** *Principles and Practice of Information Theory.* Reading, MA: Addison-Wesley, 1987, p. 58–64.
6. **Breiman, L.** The individual ergodic theorem of information theory. *Ann. Math. Stat.* 28: 809–811, 1957.
7. **Broomhead, D. S., and G. P. King.** Extracting qualitative dynamics from experimental data. *Physica D* 20: 217–236, 1986.
8. **Casdagli, M.** Nonlinear prediction of chaotic time series. *Physica D* 35: 335–356, 1989.
9. **Eckmann, J. P., and D. Ruelle.** Ergodic theory of chaos and strange attractors. *Rev. Mod. Phys.* 57: 617–656, 1985.

10. **Farmer, J. D., and J. J. Sidorowich.** Predicting chaotic time series. *Phys. Rev. Lett.* 59: 845–848, 1987.
11. **Fleming, W. H., and R. Rishel.** *Deterministic and Stochastic Optimal Control.* Berlin: Springer-Verlag, 1975.
12. **Fraser, A. M., and H. L. Swinney.** Independent coordinates for strange attractors from mutual information. *Phys. Rev. A* 33: 1134–1140, 1986.
13. **Goldberger, A. L., and D. R. Rigney.** Nonlinear dynamics at the bedside. In: *Theory of Heart*, edited by L. Glass, P. Hunter, and A. McCulloch. New York: Springer-Verlag, 1991, p. 583–605.
14. **Goldberger, A. L., D. R. Rigney, J. Mietus, E. M. Antman, and J. Greenwald.** Nonlinear dynamics in sudden cardiac death syndrome: heart rate oscillations and bifurcations. *Experentia Basel* 44: 983–987, 1988.
15. **Goldberger, A. L., and B. J. West.** Fractals in physiology and medicine. *Yale J. Biol. Med.* 60: 421–435, 1987.
16. **Grassberger, P., and I. Procaccia.** Estimation of the Kolmogorov entropy from a chaotic signal. *Phys. Rev. A* 28: 2591–2593, 1983.
17. **Grassberger, P., and I. Procaccia.** Measuring the strangeness of strange attractors. *Physica D* 9: 189–208, 1983.
18. **Gross, D., and C. M. Harris.** *Fundamentals of Queueing Theory* (2nd ed.). New York: Wiley, 1985.
19. **Jackson, J. R.** Networks of waiting lines. *Operations Res.* 5: 518–521, 1957.
20. **Kaplan, D. T., M. I. Furman, and S. M. Pincus.** Techniques for analyzing complexity in heart rate and beat-to-beat blood pressure signals. In: *Proc. IEEE Comput. Cardiol. 1990.* In press.
21. **Kaplan, D. T., M. I. Furman, S. M. Pincus, S. M. Ryan, L. A. Lipsitz, and A. L. Goldberger.** Aging and the complexity of cardiovascular dynamics. *Biophys. J.* 59: 945–949, 1991.
22. **Kleiger, R. R., J. P. Miller, J. T. Bigger, and A. J. Moss.** Decreased heart rate variability and its association with increased mortality after acute myocardial infarction. *Am. J. Cardiol.* 59: 256–262, 1987.
23. **Kluge, K. A., R. M. Harper, V. L. Schechtman, A. J. Wilson, H. J. Hoffman, and D. P. Southall.** Spectral analysis assessment of respiratory sinus arrhythmia in normal infants and infants who subsequently died of Sudden Infant Death Syndrome. *Pediatr. Res.* 24: 677–682, 1988.
24. **Kolmogorov, A. N.** A new metric invariant of transient dynamical systems and automorphisms in Lebesgue spaces. *Dokl. Akad. Nauk SSSR* 119: 861–864, 1958.
25. **Lee, A. J.** *U Statistics: Theory and Practice.* New York: Marcel Dekker, 1990.
26. **Lombardi, F., G. Sandrone, S. Pernpruner, R. Sala, M. Garimoldi, S. Cerutti, G. Baselli, M. Pagani, and A. Malliani.** Heart rate variability as an index of sympathovagal interaction after acute myocardial infarction. *Am. J. Cardiol.* 60: 1239–1245, 1987.
27. **Mayer-Kress, G., F. E. Yates, L. Benton, M. Keidel, W. Tirsch, S. J. Poppl, and K. Geist.** Dimensional analysis of nonlinear oscillations in brain, heart, and muscle. *Math. Biosci.* 90: 155–182, 1988.
28. **Ornstein, D. S.** Bernoulli shifts with the same entropy are isomorphic. *Adv. Math.* 4: 337–352, 1970.
29. **Ornstein, D. S., and B. Weiss.** How sampling reveals a process. *Ann. Prob.* 18: 905–930, 1990.
30. **Parer, W. J., J. T. Parer, R. H. Holbrook, and B. S. B. Block.** Validity of mathematical methods of quantitating fetal heart rate variability. *Am. J. Obstet. Gynecol.* 153: 402–409, 1985.
31. **Peligrad, M.** On the central limit theorem for r-mixing sequences of random variables. *Ann. Prob.* 15: 1387–1394, 1987.
32. **Pincus, S. M.** A class of Bernoulli random matrices with continuous singular stationary measures. *Ann. Prob.* 11: 931–938, 1983.
33. **Pincus, S. M.** Approximate entropy as a measure of system complexity. *Proc. Natl. Acad. Sci. USA* 88: 2297–2301, 1991.
34. **Pincus, S. M.** Approximate entropy: a complexity measure for biologic time-series data. In: *The Proceedings of the IEEE 17th Annual Northeast Bioengineering Conference*, edited by M. Fox, M. A. F. Epstein, R. B. Davis, and T. Alward. New York: IEEE Press, 1991, p. 35–36.
35. **Pincus, S. M.** Approximating Markov chains. *Proc. Natl. Acad. Sci. USA* 89: 4432–4436, 1992.
36. **Pincus, S. M., T. R. Cummins, and G. G. Haddad.** Heart rate control in normal and aborted-SIDS infants. *Am. J. Physiol.* 264 (*Regulatory Integrative Comp. Physiol.* 33): R638–R646, 1993.
37. **Pincus, S. M., I. M. Gladstone, and R. A. Ehrenkranz.** A regularity statistic for medical data analysis. *J. Clin. Monit.* 7: 335–345, 1991.
38. **Pincus, S. M., and W. M. Huang.** Approximate entropy: statistical properties and applications. *Commun. Statist. Theory Meth.* 21: 3061–3077, 1992.
39. **Pincus, S. M., and D. L. Keefe.** Quantification of hormone pulsatility via an approximate entropy algorithm. *Am. J. Physiol.* 262 (*Endocrinol. Metab.* 25): E741–E754, 1992.
40. **Pincus, S. M., and R. R. Viscarello.** Approximate entropy: a regularity measure for fetal heart rate analysis. *Obstet. Gynecol.* 79: 249–255, 1992.
41. **Shaw, R.** Strange attractors, chaotic behaviour, and information flow. *Z. Naturforsch. A* 36: 80–112, 1981.
42. **Singer, D. H., G. J. Martin, N. Magid, J. S. Weiss, J. W. Schaad, R. Kehoe, T. Zheutlin, D. J. Fintel, A.-M. Hsieh, and M. Lesch.** Low heart rate variability and sudden cardiac death. *J. Electrocardiol.* 21: S46–S55, 1988.
43. **Sugihara, G., and R. M. May.** Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature Lond.* 344: 734–741, 1990.
44. **Takens, F.** Invariants related to dimension and entropy. In: *Atas do 13. Col. Brasiliero de Matematicas.* Rio de Janeiro, Brazil: 1983.
45. **Whitt, W.** Planning queueing simulations. *Management Sci.* 35: 1341–1366, 1989.
46. **Wolf, A., J. B. Swift, H. L. Swinney, and J. A. Vastano.** Determining Lyapunov exponents from a time-series. *Physica D* 16: 285–317, 1985.
47. **Ziv, J.** On classification with empirically observed statistics and universal data compression. *IEEE Trans. Inform. Theory* 34: 278–286, 1988.